MEASUREMENT, DESIGN AND STATISTICAL METHODS IN BEHAVIOURAL RESEARCH.

A LECTURE

Delivered

By

ADELEKE, JOSHUA OLUWATOYIN (PHD.)

Institute of Education, University of Ibadan, Ibadan, Nigeria, West Africa.

Tuesday, 28 February, 2017

MEASUREMENT, DESIGN AND STATISTICAL METHODS IN BEHAVIOURAL RESEARCH.

A LECTURE

Delivered

By

ADELEKE, JOSHUA OLUWATOYIN (PHD.)

Institute of Education, University of Ibadan, Ibadan, Nigeria, West Africa.

Tuesday, 28 February, 2017

Institute of Education, University of Ibadan,
Ibadan, Nigeria, West Africa

First published 2017

New impression 2019

ISBN 978-978-950-328-6

Printed in Nigeria by
Fasco Printing Works
67 Gbadebo Street Mokola, Ibadan
08032934309

A Staff Lecture delivered
On 28 February, 2017.

Organised by

Links programme and Outreach Services
Unit of the Institute of Education,
University of Ibadan

The Director, Institute of Education,
The Deans of Faculties, Postgraduate School and Students,
HODs and Heads of Units
Special Guests, invited guests, gentlemen of the press,
Ladies and Gentlemen

I consider the opportunity to deliver this lecture as a special honour and privilege. The last lecture of this status was delivered by Dr C.V. Abe, that is, lectures presented on behalf of the Institute of Education (IoE) by research fellows of the Institute. IoE, being a research centre, has been committed to the training of research minded people on the best practices in the conduct of research, both self-initiated and directed research. IoE's services will ever be on demand because no educational system runs well without being driven by outcomes of valuable and goal oriented research. This lecture aims at revealing the essential pathways for the execution of behavioural research. On this note, I welcome everyone to this special lecture.

Nature has placed the need for research on human beings. You will expect a set of twins, born on the same day and who accessed equal opportunities from birth to have equal weight, height, complexion, cognitive ability and so on. What is commonly found are variations in those attributes. Some factors must be responsible for such differences which may attract the attention of curious minds. Since the measure of an attribute changes from individual to individual, then variation exists and the phenomenon being described is a variable. This is a convenient point to begin this discourse.

Variable

What attracts an investigator to settle for a research work is the varied nature of phenomena. According to Spiegel and Stephens (1999), a variable is a symbol, such as X, Y, H, or B that can assume any prescribed set of values called the domain of the variable. If the variable can assume only one value, it is called a constant. Variables aren't always 'quantitative' or 'numerical' otherwise they will not connote desired meanings. The variable 'gender' consists of two text values: 'male' and 'female'. One can, if it is useful, assign quantitative values instead of the text values, but we don't have to assign numbers in order for something to be a variable. It's also important to realize that variables aren't only things that we measure in the traditional sense. For instance, in most social research and in programme evaluation, we consider the treatment or programme to be made up of one or more variables (i.e., the 'cause' can be considered a variable). An educational programme can have varying amounts of 'time on task', 'classroom settings', 'student-teacher ratio', and so on. Even the program can be considered a variable (which can be made up of a number of sub-variables often referred to as treatment levels). A master degree student who opted for an experimental study for her master degree project once came to me with well coded and complete data. The data collected were actually adequate but she never knew how to derive the variable called 'treatment' because there was no item existing on any of her instruments to measure it. Such dilemma was just overcome by introducing this new variable named 'treatment' and which enabled values to be assigned nominally to the groups. That was just the solution to her problem. A researcher needs to note that any variable under investigation has well-defined measures which ensure that the domain of such variable has exhaustive and exclusive attributes. A researcher will do well to ensure that the afore mentioned features are given special attention while measuring any phenomenon.

Completely Exhaustive Levels

As rightly explained, a *variable* is something that can change, such as 'academic qualification' if at all it is typically the focus of a study.

Levels are sub-units of a variable, such as 'NCE', 'HND' First Degree, etc. An *exhaustive* list contains all possible answers. An attitudinal item may have an exhaustive list as follows:1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree. The role a variable plays within the context of a study will determine how it should be measured. For example, if age plays the role of independent variable in a study, the best approach to measure it should be by ordinal scale: 12-15 years, 16-19 years etc. and not by Interval Scale: 12, 13, 14, 15, 16, 17, 18 etc. However, if parametric statistical tool like ANOVA is to be used and age serves as dependent variable, the approach to measuring this variable will be the other way round i.e measured by interval scale.

***Mutually Exclusive* Attribute**

It is expected that no two levels of a variable are expected to occur at the same time. Thus in a survey, a person may be requested to select one answer from a list of alternatives (as opposed to selecting as many that might apply to the respondent). A person who has both NCE and First degree certificates will check both attributes on the list. Best approach to handle this is to ask the respondent to "check all that apply" and this will imply listing a series of categories? Technically speaking, each of the categories in a question like that is its own variable and is treated dichotomously as either "checked" or "unchecked", attributes that *are* mutually exclusive. While this might seem obvious, it is often rather tricky in practice. For instance, you might be tempted to represent the variable "Academic Qualification" with levels like "O Level" "NCE" "OND" "HND" "First Degree" "Master" etc. But these attributes are not necessarily mutually exclusive. In case the investigator intends to use the levels of the variable to categorise the participants,the way out is to redefine the variable to allow for one option response as against multiple option response. The new variable name will be 'Highest Academic

Qualification'.

A researcher ought to ensure that the two major attributes of a variable as earlier discussed are given considerations at instrument development stage. For instance, if the variable is "Marital Status" and the available options are "Single", "Married", and "Divorced", there are quite a few categories one can think of, that haven't been included. The list has not exhausted all possibilities. An attempt to provide an exhaustive list will pose a challenge to a researcher, especially now that there are many definitions of marital status. The way to deal with this is to explicitly list the most common attributes and then use a general category like "Others" to account for all the remaining levels. In addition to being exhaustive, the attributes of a variable should be ***mutually exclusive***, as no respondent should have two levels matching his/her status.

Redundant Sub-Level

In making levels of a variable completely exhaustive, it is a common practice of many researchers to include terms like 'undecided' or 'Not Applicable' among the levels of affective measures e.g attitude, interest, anxiety etc. Such additional level may not be needful all the time, hence it is redundant. Consider these scenarios:

1. Suppose an investigator is interested in investigating the attitude of students (who all offer Mathematics) to their Mathematics teacher. He decides to adopt the Likert response format. 1 = strongly disagree, 2 = disagree, 3 = undecided 4 = agree, 5 = strongly agree.
2. A health worker approached the same set of students to seek their disposition to the relocation of an abattoir. She opted for the same response format for the attitude items

The inclusion of 'undecided' is needless in the first scenario, since every student is expected to have attitude to his/her mathematics teacher whether positive or negative, whereas it is highly needful in the second scenario since the relocation of the abattoir is not to be their concern directly in any way.

Status of a Variable

Another important issue is the distinction between an ***independent*** and ***dependent*** variable. This distinction is particularly relevant when you are investigating cause-effect relationships. An **independent variable** is the variable manipulated by the researcher while the dependent variable is the variable of investigation i.e. what is affected by the independent variable. For example, if a researcher is studying the effect of a new Teaching Strategy on students' achievement in Chemistry, the teaching strategy is the independent variable and your measure of achievement is the dependent variable. It is worth stressing that a variable cannot assume the two status (***independent*** and ***dependent) within the context of a study.***

Forms of Variable

A variable can take various forms. The knowledge of the form(s) a variable takes will support a researcher on the choice of scale to measure it, the type of research hypothesis or research question to raise and specific statistical method that will be appropriate for the analysis.

Dependent variable is the variable of investigation.
Independent variable is the variable manipulated by the researcher
Attribute variable: is a personal characteristic/trait of the object under investigation
Alterable variable: is a variable an investigator can subject to systematic manipulation to cause effect on another variable called dependent.

Moderator: is a nuisance variable built into the study to control its inherent effect.

Descriptive variable: is a variable that will be reported on without relating it to anything in particular.

Categorical variable: results from a selection from categories, such as 'agree' and 'disagree'. Nominal and ordinal variables are categorical.

Numeric variable: gives a number, such as age, achievement score, income, etc..

Discrete variable: is a numeric variable that comes from a limited set of numbers.
It may result from, answering questions such as 'number of children', 'how often do you eat daily?', etc.

Continuous variable: is a numeric variable that can take any value, such as weight, height etc.

Measurement Scales

Scales are devised for measuring variables in behavioural research. During the past few decades, thousands of scales have been designed by researchers in sociology, psychology, education, psychiatry, ethics, behavioural science, economics, administration and other fields. Any of these scales can be categorised into one of the following four classifications:-

Nominal Scale: Some data are measured at the **nominal level.** That is, numbers used are mere labels : they express no mathematical properties. Examples are Sex (1-Male, 2-Female); State of Origin (1-Abia, 2-Adamawa,……..,36- Zamfara)

Ordinal Scale: Some data are measured at the **ordinal level**. Ordinal numbers indicate the relative position of items, but not the magnitude of difference. One example is a Likert scale:

Statement: Mathematics will not be useful to me in the future.

Response options: 1. Strongly Disagree, 2. Disagree, 3. Agree, 4. Strongly Agree

Interval Scale: Some data are measured at the **interval level**. Such numbers indicate the magnitude of difference between items, but there is no absolute zero point. Examples are days in a week and weeks in a month.

Ratio Scale: Some data are measured at the **ratio level**. Numbers at the ratio level indicate magnitude of difference and there is a fixed zero point. Ratios can be calculated. Examples include: age, income, price, costs, achievement score in a subject.

Validity of Measures

Another concept to bring to the fore before leaving the issue of measurement is what is called '*validity*'. It's best explained with an example. Adeleke, a Research Fellow in the Institute of Education, University of Ibadan, reported that his latest study showed a *significant relationship* between level Spatial ability and achievement in Geometry. Students with high spatial ability, were found to be better in Geometry than their counterparts with low Spatial ability. How did he come up with this finding? He used the best and most popular Spatial Ability test available. All his subjects were of the same age and all had the same level of education. All the students of course, spoke English and could read the test questions easily. He followed the test's rules for administration, and tested everybody together in the same classroom to make sure the testing situations were equal. He had research assistants who made sure that no one cheated. Everyone had two hours to complete the test. Adeleke collected and analysed the data obtained through test, and the result which corroborated his hypotheses showed a high level of confidence. But there's one major point here that he missed. Can you guess what it is? The test was developed in a different cultural background to the testees. The testees were not familiar with names and terms used in the Spatial Ability test, however they went ahead guessing. *Adeleke wasn't really measuring what he thought he was measuring*. He thought he was measuring Spatial Ability; instead, he succeed in obtaining data on guessing ability. His research wasn't *valid* - it didn't measure what he said it did. How could he have made his research valid? He should have replaced the strange words/terms with those ones testees are familiar with.

Basic Hints for Researchers on Measurement

1. Understand the status of a variable to be measured in your research before opting for a scale.
2. If a variable is continuous in status but is intended to be used as categorical variable, it is better to use Ratio scale rather than ordinal scale.
3. Never attempt to match variables at measurement level.
4. Use percentile ranking rather than raw scores while converting continuous measure to ordinal measure
5. Ensure complete exhaustiveness of the domain of a variable, and at the same time, avoid redundant levels.

The five basic hints enumerated above are expected to stir up questions which will be addressed at specified times. Measurement issues in behavioural research cannot be comprehensively addressed in a single lecture, however, my lecture would have thrown light into some dark areas.

Design in Behavioural Research

Imagine this scenario: Kingsley was advised to explore all the benefits that adequate designs offer a researcher. He was however warned about threats his experimental study would face if they are not identified and controlled for. He began to wonder about the sources of such threats. Can any help come to him through this lecture? It is good to begin the journey with definitions of basic terms, to throw more light on this chosen pathway called design in behavioural research.

Research design is the **plan** and **structure** of an investigation, conceived so as to obtain answers to research questions. The **plan** is the overall scheme or programme of the research. It includes an outline of what the investigator will do, from writing the hypotheses and their operational implications to the final analysis of data. A **structure** on the other hand is the framework, organization, or configuration of related elements in specified ways. Kerlinger and Lee (2000)

Sources of Error Variance

Kerlinger and Lee (2000) identified major sources of errors in a typical experimental study. Four major ones are:

Measurement: This effect occurs when measuring participants especially on an alterable phenonmenon, changes them. The influence of Post-X measure due to increased sensitization due to pretest and not by manipulation of X can be termed Measurement effect.

History: Between the Y_b (before treatment) and Y_a (post treatment) measures, many things can occur other than X (treatment). The longer the period of time, the greater the chance of extraneous variables affecting the participants, and thus the Y_a measure. This is referred to by Campbell (1957) as History effect. These variables or events are specific to the particular experimental situation.

Maturation: This covers events or variables that are generally not specific to any particular situation. They reflect change or growth in the organism studied. Mental age increases with time, an increase that can easily affect achievement, memory, and attitudes.

Statistical Regression: A statistical phenomenon that has misled researchers is the so-called regression effect. Test scores change as a statistical fact of life. On retest, test scores on the average, regress towards the mean (In statistics, **regression toward (or to) the mean is the phenomenon that if the score on a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement—and, paradoxically, if it is extreme on its second measurement, it will tend to have been closer to the average on its first**). The regression effect operates because of the imperfect correlation between the pretest and posttest scores. If r_{ab} = 1.00, then there is no regression effect; if r_{ab} = .00, the effect is at maximum in the sense that the best prediction of any posttest score is the mean. With the correlations found in practice, the net effect is that lower scores on a pretest tend to be higher, and higher scores lower on the posttest- when, in fact, no real change has taken place in the dependent variable. Thus, if low-scoring participants are used in a study, their scores on the posttest will probably be higher than on the pretest due to regression effect. This can deceive the researcher into believing that the experimental intervention has been effective when really it has not.

To lessen the effects of these possibilities, designs and statistics play major roles

The Role of Design

Kerlinger and Lee (2000) specified the three major roles and adequate design will play in a study. They are:

I. Maximization of the variance of the variable or variables of the substantive research hypothesis,

ii Control of the variance due to extraneous or unwanted variables that may have an effect on the experimental outcome, and

iii. minimization of the error or random variance, including errors of measurement, maturation and history.

Based on the above stated roles, it can then be discussed that research design is a method of variance control.

Faulty Designs

There are four (or more) inadequate designs of research that have often been used-and are still occasionally used in behavioural research. These are diagramatised below

1.

Design 1: One group			
(a)	X	Y	(Experimental)
(b)	(X)	Y	(nonexperimental

2.

Design 2: One group, Before_After (Pretest, posttest)			
(a) Y_b	X	Y_a	(Experimental)
(b) Y_b	(X)	Y_a	(nonexperimental)

3

Design 3: Simulated Before-After	
X	Y_a
Y_b	

4.

Design 4: Two groups, No control			
(a)	X	Y	(Experimental)
	~X	Y	
(b)	X	Y	(Non-Experimental)
	(~X)	Y	

A researcher may have dilemma on a chosen topic which will make the adoption of one of the adequate designs not feasible. The dilemma may border on failure to obtain pre-treatment measure and getting a comparable group to the experimental group. For example a researcher may want to use teleconferencing approach to train all the incumbent governors in Nigeria on e-governance. Assuming the approach to data collection is online. An attempt to introduce pretest measures will slow down the study because retrieving their responses may take quite a long time. Getting a comparable group may constitute another challenge, because some group of the ex-governors will not fit into the description of the population. When a researcher probably has limited time, he might opt for Design 1 in spite of its inadequacies. If obtaining pretest measures become feasible, then design 2 can be adopted. However, if the researcher has enough time at his disposal for the study so much that he can wait for another republic when a new set of governors are elected into office who will be used as simulated control group then the adoption of design 3 becomes a possibility. In a situation where the incumbent governors are randomly assigned to experimental and control groups, though pretest measurement is not feasible, Design 4 can be employed. Regardless of the dilemma a researcher is facing, any chosen design must satisfy all the essential criteria for research design.

Criteria for Research Design Selection

The choice of a design should be premised around how such design

5. Design 5: Experimental Group-Control group: Randomized participants

[R]	X	Y	(Experimental)
	~X	Y	(Control)

6 Design 6: Experimental Group-Control group: Matched participants

$[M_r]$	X	Y	(Experimental)
	~X	Y	(Control)

7. Design 7: Two groups, Before_After Randomised

(a)		Y_b	X	Y_a	(Experimental)
	[R]	Y_b	~X	Y_a	(control)
(b)		Y_b	X	Y_a	(Experimental)
	$[M_r]$	Y_b	(~X)	Y_a	(Control)

8 Design 8: Simulated Before-After; randomized

		X	Y_a	(Experimental)
[R]				
	Y_b			(Control)

9 Design 9: Three groups, Before-After

	Y_b	X	Y_a	(Experimental)
	Y_b	~X	Y_a	(control 1)
[R]		X	Y_a	(control 2)

10 Design 10: Four groups, Before-After (Solomon)

	Y_b	X	Y_a	(Experimental)
	Y_b	~X	Y_a	(control 1)
[R]		X	Y_a	(control 2)
		~X	Y_a	(control 3)

In behavioural research, perfect randomization may not be perfectly feasible from ethical point of view. If 30 students are in a class, what should make x number of students in that class qualify for a treatment and the others are denied. This explains why most of the experimental studies carried out in education adopt quasi-experimental study. Matching of the participants on certain attribute variables can strengthen systematic variance. There are so many approaches to adopt but five of the them will be presented in line with the view of Kerlinger and Lee (2000).

Matching in Behavioural Research

Although randomization which includes random selection and random assignment, is the preferred method for controlling extraneous variance, there is merit also in the use of matching. This is preferable when the universe of prospective participants is not attainable.

1. **Matching by Equating Participants**

   The most common method of matching is to equate participants on one or more variables to be controlled. Christensensen (1996) refers to this method as the **precision control method**.

2. **The Frequency Distribution Matching Method**

   The individual-by-individual matching technique presented earlier in 1 is very good for developing equal groups, but many participants must be eliminated because they cannot be matched. The frequency distribution method attempts to overcome this disadvantage while retaining some of the advantages. This technique, as its name implies, matches groups of participants in terms of overall distribution of the selected variable or variables rather than on an individual -by-individual basis.

3. **Matching by Holding Variables Constant**
   Holding the adventitious variable constant for all experimental groups is another technique that can be used to create equal groups of participants. All participants in each experimental group will have the same degree or type of extraneous variable. An example is using only male participants in both experimental and control groups to control for gender.

4. **Matching by Incorporating the Nuisance Variable Into the Research Design**
   Another way of attempting to develop equal groups is to use the nuisance or extraneous variable as an independent variable in the research design.

5. **Participants as Own Control**
   Since each individual is unique, it is difficult if not impossible to find another individual who would be a perfect match. However, a single person is always a perfect match to himself or herself. One of the more powerful techniques for achieving equality or constancy of experimental groups prior to the administration of treatment is to use that person in every condition of the experiment. The weakness of this approach is the overlap tendency of the effects of various treatments since it is difficult to make someone to unlearn what he has learnt.

Factorial Matrix

I want to submit that factorial matrix presented by a researcher can provide better understanding on how independent variables are juxtaposed to reveal possible interaction effects. The analyst will do the arrangement of both the independent and moderator variables based on the order the hypotheses are presented, at analysis level. The concept of Row-Column paradigm has been adopted in the Institute of Education for quite a long time. It is worth stressing that a researcher should keep abreast of major implications of Factorial matrix on Behavioural Research: 1) Sample size determination: a study adopting 2x2 matrix will need minimum of 4x6=24 participants; 3x3x2 will demand minimum of 18x6=108 participants and so on. 2) Inclusion of major nuisance variables that can be moderated for. A researcher who is interested in adopting factorial matrix should beware of loading the matrix with juxtaposition of many independent variables because the outcome effect may be quite unwieldy. Consequently, a researcher should note that:

- Average frequency count per cell must be equal or greater than 5
- Juxtaposition of the variables should follow Row –column-row-colum.........style.
- Involvement of more than four independent variables into the factorial matrix will make it unwieldy and produce cumbersome results consequently.

Any experimental design a researcher intends to adopt should be subjected to the underlisted test to ensure its suitability for use and its adequacy in enhancing both internal and validity of the findings.

Test of Adequacy of Randomized Subjects Designs

1. Post-Treatment measurement: Experimental time should not be too long to control for history and maturation effect.
2. Flexibility and Applicability of the design: Variables that need to be controlled can be incorporated into the design.
3. Equality of Groups: challenge around equality of groups is overcome through (a) involvement of enough participants, (b) randomization of participants and treatments, (c) Check for equality of groups on other variables other than Y. Such variables are intelligence, aptitude, or achievement.

4. Unequal number in the cells: It is expected that cells should have equal frequency counts (fs) of the participants, it is however possible to work with unequal **fs.**
5. Precise way of randomizing: Compared with matched groups designs, randomized subjects designs are usually less precise. Authorities advise that less prominence should be given to the issue of precision.

Correlated Groups

A basic principle is behind all correlated groups designs: there is systematic variance in the dependent variable measures due to the correlation between the groups on some variables related to the dependent variable. This correlation and its concomitant variance can be introduced into the measures, and the design, in three ways:

1. use the same units, for example, participants, in each of the experimental groups.
2. match units on one or more independent variables that are related to the dependent variable, and
3. use more than one group of units, like classes or schools in the design.

Multi-Group Correlated-Groups Designs

Let's suppose an investigator chooses a sample of five schools for their variety and homogeneity. The goal, of course, is to achieve external validity that is, representativeness. The investigator uses pupils from five primary schools and combines the measures from the five schools to test the mean differences in some dependent variables say Numeracy and Lifeskill. In so doing, the investigator is ignoring the variance due to the differences among schools. It is understandable that the means do not differ significantly; nevertheless, the schools variance is mixed in with the error variance. Hence Unit (such as school) variance must therefore be identified and controlled, whether it be by experimental or statistical control, or both.

Factorial Correlated-Groups Designs

Factorial models can be combined with units' notion to yield a valuable design: **Factorial correlated groups** design. Such a design is appropriate when units are natural parts of a research situation. For instance, the researcher may require the comparison of a variable before and after an experiment or intervention, or before and after an important event. Obviously, there will be correlation between the before and after dependent variable measures. Useful example is a 2X3X5 factorial design presented below.

		Treatment	
Aptitude	**School (Unit)**	**Demonstration**	**Lecture**
Low	1		
	2		
	3		
	4		
	5		
Moderate	1		
	2		
	3		
	4		
	5		
High	1		
	2		
	3		
	4		
	5		

The layout may be a bit different from what we were used to in the IoE. Since only 'change' seems to be permanent in nature, the order presented above is now the agreed position. This is a product of three independent presentations on 'factorial matrix' by Prof. Adewale, Drs. Adegoke and Adeleke to the academic staff during round table discussion sessions. **Basic Hints for Researchers**

1. **A measure of a variable should not be given different nomenclature.** An outline of a design should follow the format below:

   Experimental group 1 - O_1 X_1 O_2 O_3
   Experimental group 2 - O_1 X_2 O_2 O_3
   Control group - O_1 X_3- O_2 O_3

   Where; O_1 - represents pretest measure, O_2 - represents diagnostic tests (formative test) and O_3 represents posttest.(Summative test). It is very wrong to use different symbols to represent a measure like pretest measure.

2. Instead of having multiples of levels of a moderator variable, the number can be reduced. For example 'Low' and 'High' will suffice for a variable like Aptitude instead of adding 'Medium' especially when the sample size is small.

3. Suitability and applicability of the chosen design is an important factor. The choice of 'Solomon four design' will require perfect equality of group, which seems difficult to attain in behavioural research.

4. A factorial design will place on a researcher a demand for minimum average of 5 per cell in the matrix before valid results can be obtained.

5. Assignment of a treatment to only one unit (a school) of a domain will make the result spurious because it will be difficult to establish what variable will the systematic variance be attributed to- the domain (school effect) or treatment effect. Hence, assignment of treatment to minimum of two units will be the way out.

Using Statistics in Educational Research

Most researchers use statistics to make their conclusions. Statistics, as useful as it is, has a way of scaring even the users. This seminar intends to give a very simple, streamlined understanding of statistics. In general, statistics is used to *describe* something or to examine *differences* among groups or *relationships* among characteristics. Statisticians will use terms like *mean, median,* and *standard deviation etc.*

Mean

Mean is just a fanciful word for average. It's the sum of *all* the values or scores, divided by the *number* of people in the study or group. Most researchers that venture into experimental and ex-post facto (Non-Experimental) research have serious business with the comparison of the group means. Conscious effort ought to be directed at accuracy of measures on which the mean is estimated.

Median

The *median,* on the other hand, is nothing more than the score or value that falls closest to the *middle.* In a set of scores, half of the individual scores are higher than the median while the other half are lower. For example, if you have five numbers: 0, 0, 5, 10, 30, the *mean* or average would be 9 (0+0+5+10+30=45; 45/5=9). The *median,* however, would be 5, for there are two scores above and two scores below.

It's pretty obvious how a *mean* or an average might be used, but why do we care about the *median*? In fact, researchers use the median to give the reader more information about the mean. While the *mean* tells you what's *average*, the *median* tells you more about what's *typical*. Consider this example: The *average* performance in EVE 717 may be 60. This is partly because the lowest score is 10 and median may be 45 but the mean can be arbitrarily influenced by few outliers (

90 and above).

Standard Deviation
Another statistic that you'll see in scientific research is the *standard deviation*. It tells you how spread out the data or information is. For example, imagine that you're going to try a new teaching strategy out. Of the two strategies prescribed to you, you've been told that over time the strategies A and B have been yielding mean scores of about 75% (SD=20) and 70% (SD=2) respectively. Which of the two strategies will you consider better? In this situation Standard Deviation speaks better than the mean itself. The scores are more spread out in strategy A whereas they are relatively close in strategy B. It is this spread between scores or values that the *standard deviation* describes.

Everything described so far falls into the realm of *descriptive* statistics. They are intended to give you good information on what the subjects and the data itself look like. It provides answer to the following questions: Are the means and medians similar? Are the findings very spread out?

The intention at this moment, is to address important issues that are not commonly addressed in most statistical method classes. The reason for the overlook may be sharp criticism or argument discussing them may attract.

Differences and Relationships
The next level of statistics attempts to show *differences* between two or more groups or *relationships* between two or more different things. Suppose, for example, a researcher wants to show that a new teaching strategy is effective in reducing the number of failures. Even if the students taught with the new strategy outperformed those

taught with conventional method, the researcher needs to understand that there's always the possibility that any improvement that *does* occur is because of chance. There's also the possibility that anything that happens does so because of some factors (effects) *other than* what the researcher is studying.

Test for *Statistical Significance*

When researchers test for statistical significance, they compare different sets of values - such as learning achievement before and after using a teaching strategy - while taking into account how many people participated in the research, how dramatic their findings seem to be, and what were the characteristics of the people they compared. They then use complicated mathematical formulae to calculate *probability* values. For the new teaching strategy, this probability value will tell us how likely it is that students under the new teaching strategy performed better because the strategy did the expected, or whether they performed better simply because of chance or due to some other unknown factors other than the treatment. If the researcher finds that the probability value is low (usually less than 5%, 1% or even 1/10th of 1%), the researcher can conclude that the new strategy really does work. These probability values - called p values - represent probabilities, but they are typically represented as $p<.05$, $p<.01$, or $p<.001$, or more specifically, as $p=.023$, $p=.0067$. For example, $p<.01$ means that there is a less than 1% chance that the new teaching strategy seemed to work because of chance alone. If probabilities are low, researchers describe them as ***statistically significant.*** These are key words used while reporting. Remember: the *lower* the *p* value, the smaller the percentage, the *greater* the significance and the *less it is* likely that something might have happened just because of chance.

How good *enough is the significance level?* It depends on what's variability and experimental imprecision. This makes it difficult to distinguish real differences from random variability.

- The human brain excels at finding patterns, even from random data. Our natural inclination (especially with our own data) is to conclude that differences are real, and to minimize the contribution of random variability. Statistical rigor prevents you from making this mistake.

A Procedural Description of Analysis of Covariance

Analysis of covariance is a form of analysis of variance that tests the significance of the differences among means of experimental groups after taking into account initial differences among the groups, and the correlation of the initial measures and the dependent variable measures. That is, analysis of covariance analyses the differences between experimental groups on Y (the dependent variable) after taking into account either initial differences between the groups that is (pretest) on Y, or differences between the groups in some potential independent variable(s), X, substantially correlated with Y, the dependent variable. The measure used as a control variable- the pretest or pertinent variable- is called a covariate.

The purpose of covariates in ANCOVA is two-fold:

i. To reduce within-group variance: in ANOVA, a researcher assesses the effect of the experiment by comparing the amount of variability in the data that the experiment can explain, against the variability that it cannot explain. If we can explain some of this 'unexplained' variance (SS_R) in terms of covariates, then it is possible to reduce the error variance, allowing for more accurate assessment of the effect of the experimental manipulation (SS_M).

ii. Elimination of confounds: In any experiment, there may be unmeasured variables that can confound the results (i.e. a variable that varies systematically with the experimental manipulation). If some variables are known to influence the dependent variable being measured, then ANCOVA is ideally suited to remove the bias of these variables. Once a possible confounding variable has been identified, it can be measured and entered into the analysis as a covariate.

Post Hoc Analysis on ANCOVA Platform

To the best of my understanding, the type of post-hoc analyses on ANCOVA platform are: (i) LSD (ii) Bonferroni and (iii) Sidak. Sidak is mostly used especially for unequal group size. To demand for Scheffe post hoc analysis on ANCOVA platform will be a fallacy.

Important Hints for Researcher on Selection of Statistical Tools

The following tips are considered very relevant to those that are in the business of researching.

1. Have better understanding of parametric and non-parametric data.
2. Keep abreast of the data demand of the statistical tool(s) chosen.
3. All researchers might not be analysts but all must understand basic analytical assumptions and limitations.
4. Recognise the dynamic nature of analytical data operations and strive to catch up with the trend.
5. When confronted with any global change in analytical approach, allow the change to change your personal inclination.

Acknowledgements

'One good turn deserves another', so goes the saying. I am therefore sincerely grateful to the Institute of Education for the honour given me to present this paper. It is indeed a rare privilege and a wonderful gesture. I appreciate my noble Institute. On this note, I say a very big thank you to my amiable Director- Prof. Folajogun V. Falaye for providing a positive learning and mentoring environment. To the teacher of teachers, Emeritus Professor PAI Obanya, I appreciate your presence Sir. The Dean of the Faculty of Education, Prof. M.K.Akinsola, for taking time out to be here and Prof. A.O.U. Onuka , Head of the Links Programme and Outreach Services, the organising Unit for this event, I thank you Sirs. To all my teachers, Professors. T.W. Yoloye, C.O. Onocha, P.N. Okpala, M.A. Araromi, Adenike E. Emeke, Ifeoma M. Isiugo-Abanihe, Drs. J.A. Adegbile, Eugenia A. Okwilagwe and Modupe M. Osokoya, I appreciate everything you've taught me. To my mentor and mum who had retired but not tired, Dr. Georgina Obaitan, I appreciate you Ma.

I specially appreciate Professor A.O.U Onuka, Dr. Eugenia A. Okwilagwe and Mr. Babajide Aremu for editing this paper thoroughly, I thank you all. To my brother and teacher, Prof. Gbenga Adewale, I have had so much fun in teaching and learning with you. Thank you for your guidance. I appreciate all other Research Fellows, The Secretary to IoE, the non-teaching staff in the Institute, lecturers from the Faculty of Education and other staff from various units of the university. To the special guests, invited from far and near, I thank you all for coming. Our students in the IoE, thank you for helping us to learn better. Remember EPO's song, 'There are no teachers, there are no students, we are all learning to learn better'.

To my darling, amiable, and dynamic wife, Adenike, I really appreciate you, particularly for your understanding and support at all times. To my children, Shalom, Shammah and Rehoboth, you are really blessed.

In all, I give God the Father, The Son and The Holy Spirit all the glory for lifting up my head. To you is all the honour.

God bless you all.

References

Adeleke, J.O. (2009) *Basics of Research and Evaluation Tool.* Somerest Publishers, Ikeja, Lagos. 213 pages. ISBN: 978-978-49876-4-6.

Box, G.E. P. , Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design and data analysis.* NY: John Wiley. General introduction.

Brown, Steven R. and Lawrence E. Melamed (1990). *Experimental design and analysis*. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 74. Discusses alternative ANOVA designs and related coefficients.

Campbell, D. (1957). Factors relevant to the validity of experiments in social settings. Psychological Bulletin, 54, 297-312.

Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research, Chicago: Rand McNally

Christensen, L.B. (1996). Experimental methodology (6th ed.) Needham Heights, Ma: Allyn & Bacon

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* . Second ed., Hillsdale, NJ: Erlbaum.

Cortina, Jose M. and Hossein Nouri (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 129.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ:Prentice Hall.

Field A. (2016) Analysis of Covariance (ANCOVA) Discovering Statistics
*www.statisticshell.com/docs/**ancova**.pdf*

Gardner, P. L. (1975). Scales and statistics. *Review of Educational Research*. 45: 43-57. Discusses assumptions of the t-test.

Garson, G.D. (2008) Online Text and Notes in Statistics for Economists. www.economicsnetwork.ac.uk/teaching

Girden, Ellen R. (1992). *ANOVA Repeated Measures*. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 84.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Quantitative Applications in the Social Sciences Series No. 118. Thousand Oaks, CA: Sage Publications.

Kerlinger F.N. & Lee H.B. (2000) Foundations of Behavioral Research. 4th Edition, Wadsworth Publisher.

McCall, R.B. ( ) Fundamental Statistics for Psychology, Harcourt Brace Jovanovich. Inc.

Moore, D. S. (1995). *The basic practice of statistics*. NY: Freeman and Co.

Spiegel, M.R. & Stephens L.J. (1999) Theory and Problems of Statistics, 3rd Edition, McGraw-Hill, Inc.